# Week 1 - ML strategy (1)

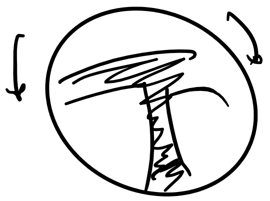
13 June 2020 11:14

## Introduction to ML strategy

### L2. Orthogonalization

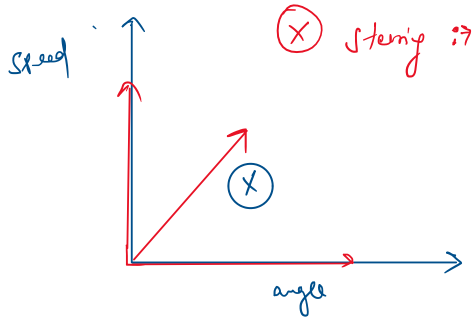
* Tuning one property at a time.

ex. driving a car.



- direction :→ steering
- Acceleration, • Braking

All have separate functions to tune




X steering :→ (0.8 * acc) + (0.3 * angle) + (0.9 * speed)

### Chain of Assumptions in ML

* fit training set well on cost function.

↓

test set

↓

dev set ✓

## Setting up your goal

### L1. Single Number evaluation metric

| Classifier | Precision | Recall |
|---|---|---|
| A | 95% | 90% |
| B | 98% | 85% |

Out of examples recognized, what % are actually cats.

What % of actual cats are correctly recognized.

## What to choose?

**F1-score** "Avg. of Precision and recall"

$$\left( \frac{2}{1/P + 1/R} \right)$$ :. Harmonic Mean

## L2. Satisficing and optimizing metrics.

| Classifier | Accuracy | Running Time |
|---|---|---|
| A | 90 % | 80 ms. |
| B | 92% | 95 ms. |
| C | 95% | 1500 ms. |

- maximize A
- minimize Running Time.

N metrics :- 1 optimizing metric.

(N-1) satisficing metrics.

## L3. Train / dev / Test distributions.

### Cat Classifier.

Regions :->

- India
- Mexico
- Canada
- Spain
- Japan
- China

dev

test

(X).

# Dev Test and test set should come from same distribution.

## L4- Size of Dev/test Sets

Train :- 70%  test :- 30%

or

Train :- 60%  dev/cv :- 20%  test :- 20%.

for 1,000,000 examples —


## Size of Test Set

\# Set your test size big enough to give high confidence in the overall performance of System.

## L5. When to change dev/test sets and metrics

Metric :- Classification error.

- Algorithm A :- 3% error → Pornographic Images
- Algorithm B :- 5% error. → ! P — Images.

Missclassification.

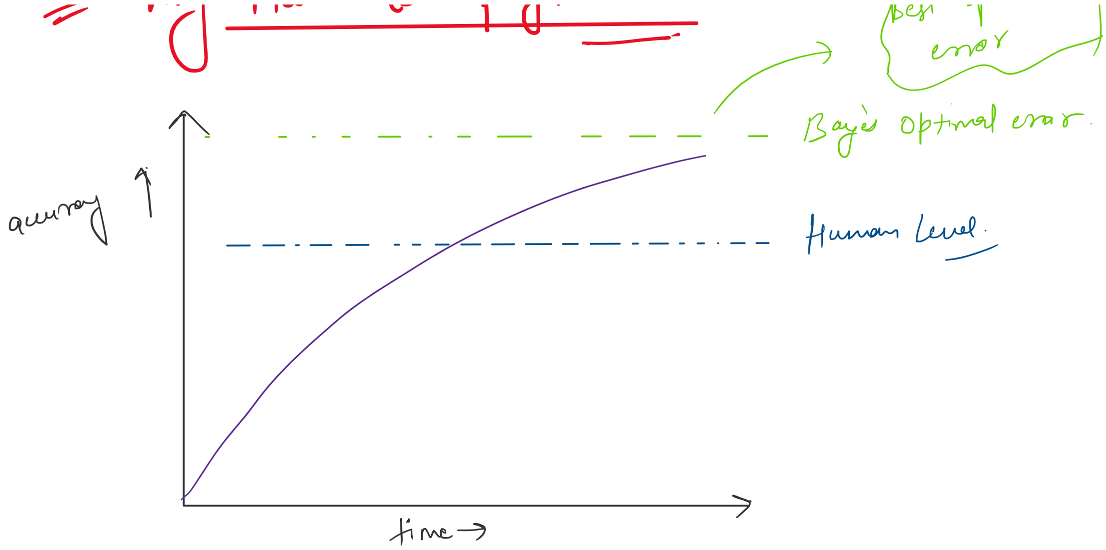
- Error :- $\frac{1}{\sum w^i} \sum_{i=1}^{m_{dev}} w^i \, I\{ y_{pred}^{(i)} \neq y^{(i)} \}$

$$w^i = \begin{cases} 1, & \text{if related content} \\ 100, & \text{if misclassified.} \end{cases}$$

Comparing to human level performance

## L1: Why Human level performance

Best optimal error




## Avoidable Bias

| | | |
|---|---|---|
| Human Error :→ | 1% | 7.5% (say) |
| Training error :→ | 8% | 8% |
| dev Error :→ | 10% | 10% |
| | * focus on bias | * focus on variance |

Avoidable bias

\# Human level error is a proxy for bayes error.

almost same content as in previous course or Machine Learning by Andrew ng.

Week 2 - ML strategy (2)

13 June 2020 19:51

## Error Analysis

### L1. Carrying Out error Analysis.

Cat classifier - 5% error / 95% accuracy.

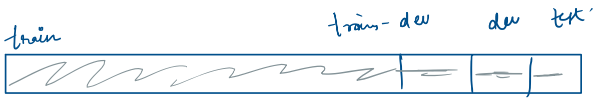
- Get 100 ~ mislabelled dev set examples.
- Count up how many are dogs.
- Count errors.

### Bias and Variance with mismatched data distribution

Assume Human got ≈ 1% error.

Training error :- 1%

Dev Error :- 10%




Variance Error.

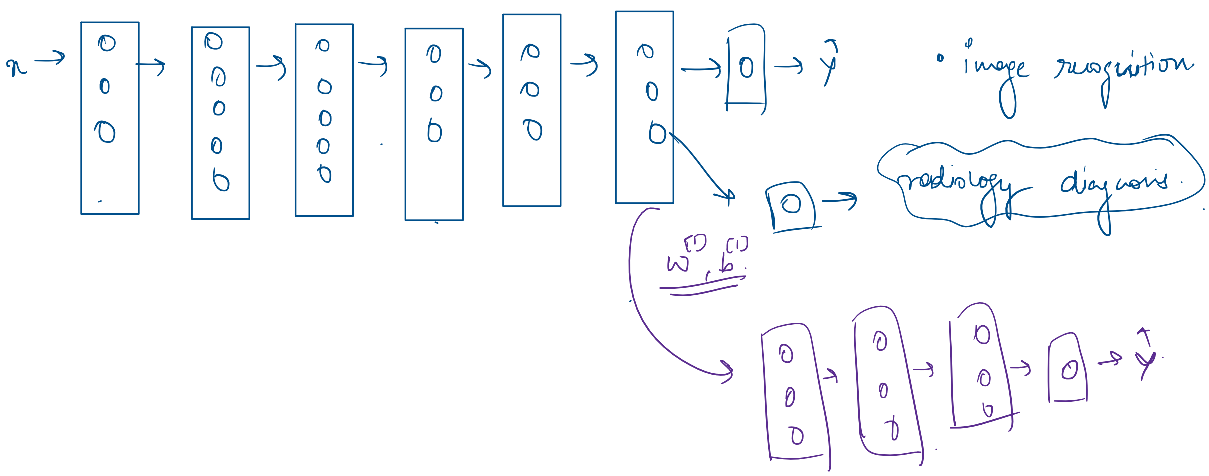
# Explain almost same as train/test/dev split in 2nd course.

### data mismatch problem:

- Carry out manual analysis and try to understand the diffrence b/w training and dev/test sets.
- Make train set more similar.

### Learn from Multiple Task.

Transfer Learning :->



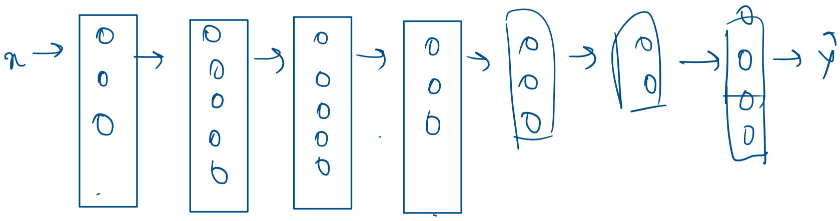

When transfer learning makes sense :->

## When transfer learning makes sense :⇒

Transfer(A) → (B) { They have same input $x$, A has more data than B. }

## Multi-Task Learning.

* Start off simultaneously




Loss $\hat{y}^{(i)} = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{4} \mathcal{L}(\hat{y}_j^{(i)}, y_j^{(i)})$

(Logistic loss.)*

* Train a set of task that could benefit from having shared low level features.

## End-to-End deep learning

Speech Recognition (ex)

X → features → phonemes → words → y

(audio)   MFCC   (transcript).

end-to-end deep learning →

## Whether to use End-to-end deep learning

Pros

- Let the data speak.
- Less hand design components needed.

Cons

- Needs large amount of data.

  X → Y
- Excludes potentially useful hand design components.